3Я

УДК 004.934

В.В. Яценко
Міжнародний науково-навчальний центр інформаційних технологій та систем,
м. Київ, Україна
val-yatsenko@uasoiro.org.ua

Параметризація типів речень предметної області для системи усного фразника-перекладача

У статті розглядаються підходи побудови системи перекладу усномовного сигналу в межах предметних областей. Блок інтерпретації отримує вимовлене речення у вигляді розпізнаної декодером послідовності слів. На виході системи приймається рішення про належність розпізнаної послідовності слів до типу речення, що задає тип смислу. Розпізнавання виконується з урахуванням параметрів, що описують множину можливих варіантів висловів у фразах при спілкуванні. Проаналізовані альтернативні підходи, щодо моделювання обмежень на допустимі послідовності слів у фразах. Описано ймовірнісне оцінювання належності розпізнаної послідовності слів до певного типу речення.

Вступ

Задача розпізнавання усномовного сигналу мови передбачає не лише автоматичне відновлення текстів, що вимовляються людиною у вигляді слів, фраз, речень природною мовою. Не менш важливо розуміти, що сказано диктором. З одного боку, це дає змогу формувати більш адекватну відповідь розпізнавання. З іншого боку, маючи змогу оперувати смислами, легко перейти до того самого смислу, але вже іншою мовою, отримуючи таким чином систему усного перекладу.

Серед важливих практичних задач, пов'язаних з розпізнаванням мови, до яких відносяться система диктування текстів, довідкова система, система мовленнєвого керування різними пристроями, системи мовленнєвого діалогу (наприклад, по телефону) тощо, ми виокремимо систему усного перекладу. Її актуальність полягає, зокрема, в автоматизації розмовника, відомого всім у вигляді паперової книжки. Ця книжка розбита на теми, має безліч сторінок. Користувач вимушений шукати потрібну фразу та ще й «озвучувати» переклад іншою мовою своїм голосом. Натомість пропонується користувачеві лише вимовити фразу рідною мовою з обраної теми. Далі система все робить сама. Додаткової уваги тут потребує питання моделювання параметрів у реченнях. Тобто необхідно передбачити всі можливі варіанти значень для певного речення. Наприклад, у питанні щодо подорожі до певного міста параметром є назва міста, оскільки існує деяка скінченна множина міст, куди можна поїхати.

Такі системи є актуальними з огляду на використання їх при розмові з іншомовною особою. Користувач отримує переклад фрази іншою мовою у візуальній або звуковій формі, що має спростити спілкування в іншомовному середовищі.

При побудові систем усного перекладу в межах предметних областей виникає ряд проблем, спільних з проблемами задачі розуміння мовленнєвого сигналу. Необхідно побудувати моделі всіх можливих речень мови діалогу, що виражають один і той самий зміст, змоделювати параметри слів у типах речень, згенерувати та відшукати найбільш правдоподібні еталонні сигнали, враховуючи параметри. Хоч і не так явно, але у слов'янських мовах все ж існують обмеження на порядок слідування слів, які слід враховувати, використовуючи лінгвістичні знання.

Для дослідження та специфікації обмежень на допустимі послідовності слів у фразах використовувалися LISP-структури [1], [2]. На основі цих структур генерується величезна кількість речень, що мають один і той самий зміст з точністю до параметрів. Втім, існує ряд обмежень на застосування цієї технології, пов'язаних як з суб'єктивним чинником при побудові LISP-структур, так і зі збільшенням обчислень, викликаних значним ускладненням графа розпізнавання.

Як альтернатива до LISP-структур пропонувався спосіб оцінювання належності послідовності слів типам речень, що характеризують смисл [3]. Цей підхід потребує розвинення, зокрема, з метою аналізу можливих помилок розпізнавання в процесі смислової інтерпретації.

Для моделювання обмежень на порядок слідування слів застосовувалися граматичні знання [2], що і надалі використовуються в дослідженнях.

Для моделювання параметрів у типах речень використовувалися бази даних та бази знань. Була сформульована лінгвістична модель інтерпретації розпізнаного сигналу з врахуванням параметрів.

Спочатку розглянуто загальну структуру системи усного перекладу, наступний розділ присвячено опису моделювання типів речень з урахуванням параметрів, далі пропонується розглянути спосіб оцінювання належності послідовності слів до типу речень, насамкінець описуються експериментальні дослідження.

Загальна структура системи усного перекладу в межах предметних областей

Розпізнавання та змістовна інтерпретація злитого мовлення виконується в єдиному взаємопов'язаному процесі. Кінцевою метою цього процесу є зміст повідомлення, який передається послідовністю слів, та його переклад іншою мовою.

Розглянемо, в чому полягають і як взаємопов'язані задачі розпізнавання та інтерпретації злитого мовлення [1], [2]. Розпізнавання мови – це процес автоматичної обробки сигналу з метою визначення послідовності слів, які передаються цим сигналом. Змістовна інтерпретація мови – це процес автоматичної обробки мовленнєвого сигналу з метою виявлення змісту, що передається сигналом, та представлення цього змісту в певній канонічній формі, зручній для подальшого використання в системі усного перекладу.

Очевидно, що змістовна інтерпретація мови є більш високим ступенем узагальнення інформації, ніж розпізнавання. Оскільки кожну думку можна висловити різними реченнями в мові діалогу, але при цьому зміст не зміниться, то слід визначити певні обмеження на допустимі послідовності слідування слів у реченнях. Тому, при інтерпретації змісту мови різні речення, що передають одну і ту саму думку, повинні відображатися в один і той же результат, тобто відповідь розпізнавання (послідовність слів) не повинна суперечити синтаксису, семантиці та прагматиці предметної області.

Зважаючи на це, пропонується розглянути структуру системи усного перекладу в межах предметних областей (рис. 1). Задача змістовної інтерпретації злитого мовлення з метою подальшого перекладу ґрунтується на тому, що спочатку користувач має задати предметну область (ПО), з якою він бажає працювати. Для цього потрібно назвати цю ПО. Взагалі розглядається 15 ПО, з якими може працювати користувач. Активатор вибирає названу з 15 ПО і завантажує підсловник ПО з відповідними до цієї області ТР та граматику, за якою моделюються допустимі обмеження на послідовності слів у реченнях.

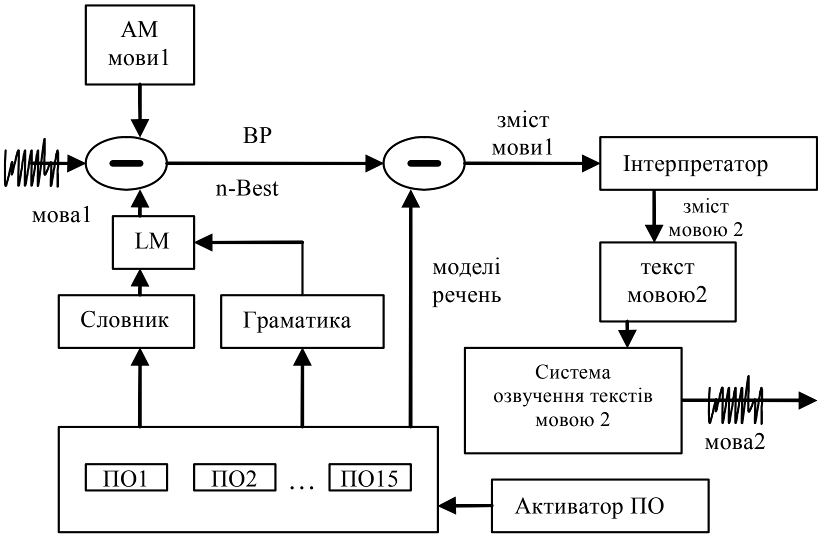


Рисунок 1 – Структура системи усного перекладу в межах предметних областей

Диктор вимовляє мовою 1 деяке речення, що розпізнається з урахуванням акустичної моделі та побудованої згідно зі словником відповідної ПО та граматики, лінгвістичної моделі (LM). Потім обирається n кращих послідовностей слів і порівнюється зі згенерованими моделями речень, які можуть задавати відповідний ТР. Використовуючи імовірнісне оцінювання, приймається рішення про належність розпізнаної послідовності слів до ТР. За цим ТР визначається ТЗ і інтерпретатор знаходить відповідний ТЗ іншою мовою. На виході ми маємо отримати текст мовою 2, який може бути озвученим відповідною системою озвучення текстів мовою 2.

В описаній структурі перекладу залишається досить складною задача інтерпретації відповіді розпізнавання. Шляхи розв’язку цієї задачі, описані в [2], ґрунтуються на економному задаванні всіх можливих допустимих речень у мові діалогу.

Таким чином, автоматичний переклад фрази вимовленої мовою 1 на мову 2 з озвученням результату, за допомогою пропонованої структури усного перекладу в рамках предметних областей буде полягати в тому, щоб спочатку для сигналу, що подається (вимовляється диктором), знайти найбільш правдоподібний, можливо, параметризований, тип речення серед всіх типів речень, які задають тип змісту, а потім визначити сам тип змісту того змістовного висловлювання, що було вимовлене, та знайти для нього відповідний тип змісту в мові 2 з урахуванням параметрів. Насамкінець, речення, отримане мовою 2, озвучується.

Далі буде розглянуто способи генерування речень мови діалогу, що виражають один і той самий зміст, але набувають певної деталізації, яка передається параметрами, та пошуку найбільш правдоподібних типів речень та розроблення обмежень на допустимі послідовності слів згідно зі структурами, якими можна представити речення.

Моделювання типів речень з урахуванням параметрів

Оскільки структура перекладу повинна працювати в межах предметних областей, то пропонується розглянути певну ієрархію мовленнєвих образів [2]. Мається на увазі, що вся діяльність людини розбивається на предметні області (ПО) за типом тем паперового розмовника. Кожна ПО складається зі скінченної множини типів змістів (ТЗ). В кожен ТЗ входить множина еквівалентно змістовних типів речень

(ТР). Тип речення – це конструкція, що економно задає множину речень, отриманих з одного речення незалежними допустимою заміною та допустимою перестановкою чи випадінням слів та словосполучень.

В межах задачі розпізнавання, інтерпретації та перекладу мовленнєвого сигналу надзвичайно важливим є питання опису параметрів слів у фразах, де можуть бути різні варіанти власних назв, часу, адрес тощо. Значення терміна «параметр» може мати різну інтерпретацію в залежності від того, в якому контексті він вживається. Загалом параметром називають величину, значення якої служить для розрізнення елементів деякої множини між собою.

У даній статті пропонується розглянути параметри, що описують множину можливих варіантів висловів у фразах при спілкуванні. Розглянемо приклад ТР прохання розбудити людину в певний час з ПО «Готель». Базова структура матиме вигляд (рис. 2):

$$\left(\left(\text{розбудіть}\right)\begin{pmatrix}\text{мене}\\ \text{нас}\\ *\end{pmatrix}\begin{pmatrix}\text{будь ласка}\\ *\end{pmatrix}\begin{pmatrix}\text{о \$time:app}\\ \begin{bmatrix}\text{в}\\ \text{у}\\ \text{через}\end{bmatrix}\text{\$time}\end{pmatrix}\right)$$

Рисунок 2 – Базова структура ТР

В дужках () вказані підсловники, які можна переставляти місцями, а в [] – які не можна переставляти. Переставляти підсловники можна лише всередині старших дужок. Символ * означає порожнє слово.

Цією структурою можна згенерувати досить багато речень з урахуванням параметрів. Серед них можуть бути, наприклад, такі:

Розбудіть мене, будь ласка о сьомій годині.

Будь ласка, нас розбудіть о п'ятій ранку.

О сьомій тридцять розбудіть мене.

Розбудіть мене через шість годин.

Розбудіть нас у шість.

У наведених прикладах останнє речення є розмовним, а не літературним, на практиці так висловлюються, і ми маємо це врахувати.

У прикладі типу речення параметром є «часові призначення»: $time:app, $time. Розглянемо перший параметр $time:app. Він описує будь-який час з точністю до, наприклад, хвилин, в контексті призначення деякої події. Цей параметр є дуже складним, оскільки людина може висловити час багатьма способами: вказувати або на одиницю виміру часу (години, хвилини), період доби. Щоб передбачити всі варіанти та значення цих параметрів вводиться спеціально розроблена і описана параметрична граматика словника на основі форми Бекуса-Наура (BNF). Таку граматику можна подати у розгорнутому вигляді (табл. 1 і 2).

В наведеному прикладі базова структура задає $4!\cdot 1\cdot 3\cdot 2\cdot 4=576$ параметризованих речень, допустимих у мові усного діалогу. Якщо врахувати, що кожен параметр містить величезну кількість варіантів, кількість речень значно зросте.

Всі речення мови діалогу можна задавати за допомогою ТС і відповідних їм ТР, використовуючи структуру, наведену у прикладі. За допомогою LISP-структур генерується величезна кількість речень, що мають один і той самий зміст. Оскільки побудова LISP-структур є досить громіздкою, потребує багато ручної роботи, то було розроблено автоматизований специфікатор предметних областей.

Таблиця 1 – Базові структури параметрів часового призначення

позначення	*приклад*	*параметризація для української мови*
$time:app	*о шостій*	$hour:nadj-at
	о шостій тридцять	$hour:nadj-at [$teen:n | $dec:5max]
	о шостій годині	$hour:nadj-at $hour-i
	о шостій годині ранку	$hour:nadj-at $hour-i $time:post
	о шостій тридцять ранку	$hour:nadj-at [$teen:n | $dec:6max] $time:post
	о шостій годині тридцять хвилин	$hour:nadj-at $hour-i $min:n-u
	о шостій годині тридцять хвилин ранку	$hour:nadj-at $hour-i $min:n-u $time:post
$min:n-u	*одна хвилина*	$min:n1 $min1
	дві хвилини; 53 хвилини	$min:n2 $min2
	5 хвилин; 37 хвилин	$min:n5 $min5
$min:n5	*20; 45*	$dec:5max [$digit5]
	5; 7	$digit5
	12; 15	$teen:n

Таблиця 2 – Опис значень параметрів для часового призначення

позначення	*укр. мова*	*англ. мова*
$hour:nadj-at	перший	one
	другий	two
	…	
	двадцять третій	twenty three
$hour-i	годині	o'clock
$min:n1	одна	one
	…	
	п’ятдесят одна	fifty one
$min:n2	дві	two
	три	three
	…	
	п’ятдесят три	fifty three
$teen:n	десять	ten
	…	...
	дев’ятнадцять	nineteen
$dec:50max	двадцять	twenty
	тридцять	thirty
	сорок	forty
	п’ядесят	fifty
$digit5	п’ять	five
	…	
	дев’ять	nine
$min1	хвилина	minute
$min2	хвилини	minute
$min5	хвилин	minute
$time:post	ранку	a.m.
	дня	p.m.
	вечора	p.m.
	ночі	a.m.

Для побудови всіх можливих речень мови усного діалогу можна використовувати так звану орієнтовану семантичну мережу (ОСМ) [1], [2]. Ця мережа одночасно задає обмежену граматику порядку слідування слів, яку можна використовувати при розпізнаванні.

Альтернативною до цієї граматики є граматика вільного поряду слідування слів. Між цими протилежними за суттю граматиками може бути побудовано безліч інших відносно вільних або відносно обмежених граматик. Ми пропонуємо дещо обмежити вільну граматику за рахунок лінгвістичного поняття про фонетичне слово.

Під фонетичним словом розуміємо слово з невіддільними від нього супутніми словами. Наприклад, невіддільними є прийменник від іменника або прикметника, частка «не» спереду дієслова і частка «б» позаду нього. Пропонована нами відносно вільна граматика подається у вигляді графа (рис. 3), де *pau* – слово-пауза на початку та в кінці фрази, *pcl* – проклітик, *prep* – прийменник, *w* – нейтральне слово, *ecl* – енклітик.

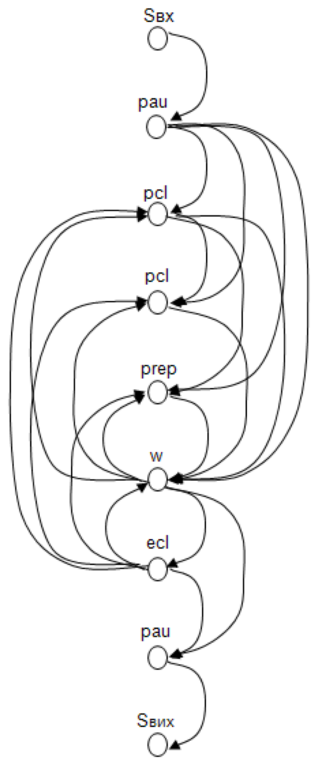


Рисунок 3 – Граф відносно вільної граматики
на основі поняття про лінгвістичне слово

Втім, за такої граматики прийняття рішення щодо смислу речення не є очевидним. Цьому питанню присвячено наступний розділ.

Статистичне оцінювання належності послідовності слів до типу речень

При розпізнаванні в умовах граматики, що не задає строгих обмежень на послідовності слів, очевидно можуть бути отримані відповіді розпізнавання, що не входять у множину речень, які згенеровані певним типом речень. Це може бути зумовлено як помилками при розпізнаванні, так і при формуванні типів речень експертом. Крім того, сам користувач може вимовити речення з різного роду відхиленнями або аграматизмами, наприклад, повторити деяке слово двічі.

Тому пропонується оцінювати ймовірність типу речення *ST* з ОСМ за умови розпізнаної послідовності слів $(w_1, w_2, ..., w_n)$ та оголошувати відповіддю інтерпретації той тип речень – *ST**, для якого ця ймовірність є найбільшою:

$$ST^* = \underset{ST}{\operatorname{argmax}}\, P(ST / w_1, w_2, ..., w_n), \qquad (1)$$

Імовірність у лівій частині (1) може бути переписана за формулою Байєса в наступному вигляді:

$$P(ST / w_1, w_2, ..., w_n) = \frac{P(ST)}{P(w_1, w_2, ..., w_n)} P(w_1, w_2, ..., w_n / ST), \qquad (2)$$

Розглядаючи послідовність $(w_1, w_2, ..., w_n)$ як марківський процес, подаємо кожний із множників умовної імовірності у правій частині (2) у вигляді:

$$P(w_1, w_2, ..., w_n / ST) = \prod_{k=1}^{n} P(w_k / ST, w_{k-m}, ..., w_{k-1}), \qquad (3)$$

$$P(w_1, w_2, ..., w_n) = \prod_{k=1}^{n} P(w_k / w_{k-m}, ..., w_{k-1}), \qquad (4)$$

де $m \geq 0$ – порядок процесу.

Оцінювання кожного з множників з правої частини виразів (3) і (4) може виконуватися різними способами в залежності від обраного порядку процесу.

Ми розглядали найпростіший випадок, коли m = 0. Тоді, враховуючи формулою Байєса, вираз (2) подаємо у вигляді:

$$P(ST / w_1, w_2, ..., w_n) = P(ST) \prod_{k=1}^{n} P(ST / w_k), \qquad (5)$$

Тут логічно зробити припущення щодо рівноімовірності всіх типів речень. Хоч насправді, деякі смисли трапляються частіше за інші, і це залежить від попереднього змісту. Залишається обчислити вираз вигляду $P(ST_k / w)$. Для цього розглянемо $ST(w_k)$ – множину типів речень, в яких зустрічається слово w_k. Тоді

$$P(ST/w_k) = \begin{cases} |ST(w_k)|^{-1}, & \text{якщо } ST(w_k) \neq \varnothing, \\ \alpha(ST, w_k), & \text{в іншому випадку.} \end{cases} \qquad (6)$$

Вираз $\alpha(ST, w_k)$ має зміст імовірності того, що слово w_k розпізнано помилково замість деякого слова w: $ST(w) \neq \varnothing$. Цю ймовірність можна оцінити на основі деякої міри мінімальної редакторської правки $d(w_k, w)$, наприклад, відстані Левенштейна. При обчисленні цієї міри штрафуються вставки, видалення та заміни символів фонемного тексту порівнюваних слів. Отже, вираз $\alpha(ST, w_k)$ пропонується оцінювати таким чином:

$$\alpha(ST, w_k) = \left(1 - \min_{ST(w) \neq \varnothing} \frac{d(w_k, w)}{L(w)}\right) ST^{-1}\left(\arg\min_{ST(w) \neq \varnothing} d(w_k, w)\right), \qquad (7)$$

де $L(w)$ – кількість фонем у слові *w*.

Рішення щодо приналежності розпізнаної послідовності слів певному типу речень приймається на основі (1) – (7).

Для апробації цього способу було проведено серію експериментальних досліджень.

Експериментальні результати

Запропоновані в роботі методи оцінювання належності послідовності слів до типу речень були експериментально перевірені на фразах, взятих із звичайного розмовника. В роботі для прикладу розглядалися три ПО: «Повсякденні фрази», «По-

дорож», «Готель». Ці ПО містять 47 + 102 + 68 = 217 типів змісту. В середньому на тип змісту припадає 4,17 базових речення.

Акустичні моделі для декодера створено на основі мовленнєвого корпусу окремо вимовлених слів, в якій брало участь 60 дикторів [2]. Засобами [3] проведено навчання 55 прихованих марківських моделей фонем. Максимальна кількість нормальних законів у суміші – 20.

Для експерименту довільним чином було вибрано 500 фраз. Смислова інтерпретація проводилася на основі відповіді пофонемного розпізнавача [4] за умов вільної та відносно вільної (на основі фонетичних слів) послівних граматик [2]. З результатів проведеного експерименту, наведених у табл. 3, випливає, що для двох типів граматик похибка змістовної інтерпретації не перевищує 5%, що є прийнятним для прикладної системи.

За умов обмеженої граматики швидкість розпізнавання в 10 разів перевищувала реальний час, а за умов вільної та відносно вільної граматик розпізнавання відбувалося швидше реального часу на ресурсах нетбука.

Таблиця 3 – Результати розпізнавання та смислової інтерпретації 500 речень з двох предметних областей

Тип граматики	Надійність розпізнавання (%)		
Обмежена	96,7	94,1	98,3
Вільна послівна	53,4	4,2	86,1
Відносно вільна	79,1	20,8	96,2

На основі проведених досліджень розроблено демонстраційну програмну модель перекладу речень, вимовлених українською мовою, на англійську мову (рис. 3). При цьому, слідування слів в українському реченні може бути будь-яким із допустимих, допускаються елементи спонтанності. Реченню, вимовленому українською мовою, ставиться у відповідність англомовний тип змісту або речення, а перше речення цього типу змісту оголошується результатом перекладу.

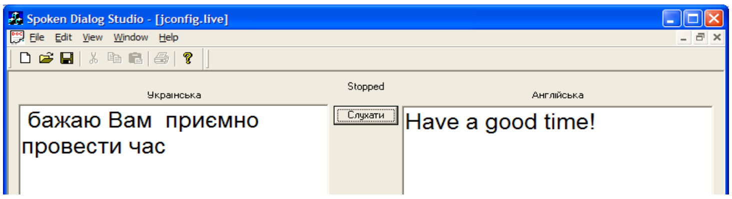


Рисунок 4 – Демонстраційне програмне забезпечення моделі усного фразника-перекладача

Висновки

Представлена в роботі система усного перекладу є електронним аналогом паперового розмовника, взаємодія з яким відбувається найбільш природними способом – голосом. При розпізнаванні вимовленої користувачем фрази використовуються лінгвістичні та семантичні знання щодо обраної предметної області. Введені при цьому «м'які» обмеження на порядок слідування слів дають змогу підвищити надійність розпізнавання, не підвищуючи вимог до обчислювальних ресурсів. Розроблені програмні засоби дають змогу формувати граматики слідування слів для розпізнавання злитого мовлення як на основі типів речень, так і на основі лінгвістичного поняття про фонетичне слово.

Використання параметричних моделей типів речень надає можливість користувачеві більш вільно та розмаїто спілкуватися, що розширює коло використання системи. Припущення, що спостережувані послідовності слів є марківським процесом, дало змогу сформулювати більш гнучкий спосіб формування відповіді смислової інтерпретації.

На основі експериментальної моделі розроблено програмну модель усного словника-перекладача для перекладу з української мови на англійську в межах предметної області, що працює в межах реального часу на обмежених обчислювальних ресурсах.

Одні і ті ж самі фрази з різною інтонацією можуть виражати як питальне речення, так і розповідне. Отже, в подальшій роботі слід дослідити можливість розпізнавання інтонації та ритму (просодики) з метою автоматичного розставляння розділових знаків у розпізнаних фразах.

Надалі також планується ставити у відповідність україномовній фразі більш точний англомовний відповідник серед типів речень з типу змісту.

Література

1. Vintsiuk T.K., Analysis, Recognition and Understanding of Speech Signals / T.K Vintsiuk. – Kyiv : Naukova Dumka, 1987. – 264 p (in Russian).
2. Sazhok M. Spoken translation system based on speech understanding in subject area / M. Sazhok, V. Yatsenko. // All-Ukrainian Int. Conference on Signal : Image Processing and Pattern Recognition UkrObraz'2010. – Kyiv, 2010. – P. 103-106 (in Ukrainian).
3. Lee T. Kawahara. Julius – an open source real-time large vocabulary recognition engine / Lee, T. Kawahara and K. Shikano // –In Proc. European Conference on Speech Communication and Technology (EUROSPEECH). – 2001. – P. 1691-1694.
4. [Young S.J.] et al., HTK Book, Version 3.1, Cambridge University. – 2002.

Literatura

1. Vintsiuk T.K. Analysis, Recognition and Understanding of Speech Signals. Kyiv: Naukova Dumka. 1987. 264 p.
2. Sazhok M. All-Ukrainian Int. Conference on Signal/Image Processing and Pattern Recognition Ukr Obraz'2010. Kyiv. 2010. P. 103-106
3. Lee T. Proc. European Conference on Speech Communication and Technology (EUROSPEECH). 2001. P 1691-1694
4. Young S.J. HTK Book, version 3.1. Cambridge University. 2002.

В.В. Яценко

Параметризация типов предложений предметной области для системы устного фразаря-переводчика

В статье рассматриваются подходы построения системы перевода устного сигнала в рамках предметных областей. Блок интерпретации получает произнесенное предложение в виде распознанной декодером последовательности слов. На выходе системы принимается решение о принадлежности распознанной последовательности слов типу предложения, которое задает тип смысла. Распознавание выполняется с учетом параметров, которые описывают множество возможных вариантов высказываний во фразах при общении. Проанализированы альтернативные подходы моделирования ограничений на допустимые последовательности слов во фразах.

V. Yatsenko

Parameterization of Sentence Types in a Subject Area for the Spoken Interpreter System

In this paper we describe the approaches for building of the spoken translation system within subject areas. The decoded sequence of words enters to the interpretation subsystem, which finally makes decision concerning to the sentence type and the respective meaning type for the pronounced utterance. Possible variations of date, time, place etc. that may occur in sentences are parameterized and integrated to the language model. Strict, free and phonetic words based on word grammars for speech decoder are analyzed.

Стаття надійшла до редакції 22.06.2011.